DOCUMENT RESUME

ED 440 365 CS 013 945

AUTHOR Gover, Mark R.; Englert, Carol Sue
TITLE Quantitative Results of the Learning Environments for Accelerated Progress (LEAP) Curriculum.
PUB DATE 1999-12-00
NOTE 19p.; Paper presented at the Annual Meeting of the National Reading Conference (49th, Orlando, FL, December 1-4, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Data Analysis; Elementary Education; Instructional Effectiveness; Multivariate Analysis; Reliability; *Scoring; Special Education; *Student Evaluation; Validity; *Writing (Composition); *Writing Evaluation; Writing Research; *Writing Skills
IDENTIFIERS Discourse Conventions; Genre Approach; *Learning Environments; Primary Trait Scoring

ABSTRACT

An overview of LEAP (Learning Environments for Accelerated Progress) project outcomes draws data from original scoring protocols developed by the LEAP project. These scoring protocols (or "measures") were developed to assess students' progress in the composition of three text genres: narrative, expert, and explanation. Initial analyses of data sought to answer two questions: (1) Are these protocols for the assessment of students' genre skills internally consistent and valid? and (2) From a statistical perspective, does the LEAP curriculum appear to have an effect? Each protocol incorporates two scales: a "primary traits" scale and a "convention" scale. The primary traits scale is intended to assess the degree to which a student captures the essential characteristics of one of the three particular genres. The writing conventions scale consisted of items intended to measure conventional writing skills such as spelling, paragraph writing, capitalizing, and so on. The sample consisted of approximately 115 special education students; although students ranged from kindergarten through sixth grade, the average was around third grade. Eight tables of data and three charts outline extensive results. Based on the evidence in the tables, both of the initial questions can be answered in the affirmative. Contains 8 tables, 3 charts, and 4 notes. (NKA)

Reproductions supplied by EDRS are the best that can be made
from the original document.

ED 440 365

Quantitative results of the Learning Environments for Accelerated Progress (LEAP) Curriculum

Mark R. Gover
Carol Sue Englert
Michigan State University

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

M. Gover

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

CS013945

BEST COPY AVAILABLE

Quantitative results of the Learning Environments for Accelerated Progress (LEAP) Curriculum

Mark R. Gover
Carol Sue Englert
Michigan State University

Paper prepared for the annual meeting of the National Reading Conference, December 1-4, 1999, Orlando, Florida

Introduction

This paper will present an initial overview, from a bird's eye view, of LEAP project outcomes. The data are drawn from original scoring protocols developed by the LEAP project. These scoring protocols (which we will refer to loosely as "measures") were developed to assess students' progress in the composition of three genre of texts: *narrative*, in which students relate a personal experience in story form; *expert*, in which students articulate their systematic conceptual knowledge about a chosen subject; and *explanation*, in which students explicate the steps required to accomplish a specific task.

Our initial analyses seek to answer two general questions. First, are these protocols for the assessment of students' genre skills internally consistent and valid? Second, from a statistical perspective, does the LEAP curriculum appear to have an effect? Analyses included tests of the internal consistency of these measures as well as comparisons of the progress made by LEAP versus comparison groups. We use the term "comparison group" in accord with a quasi-experimental design (Cook & Campbell, 1979).[1] After an overview of the original scoring protocols, we offer evidence for their consistency and validity. This will be followed by tests of the LEAP curriculum itself based on one of these protocols.

LEAP's Original Measures

The Morning News, Expert, and Explanation data are based on original scoring protocols. Among other items, each protocol incorporates two scales: a "primary traits" scale and a "convention" scale. The primary traits scales are intended to assess the degree to which a student captures the essential characteristics of one of three particular genre (Table 1).

---

[1] Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: Design and analysis issues for field settings.* Boston: Houghton Mifflin.

CS 013 945

Table 1

THREE GENRE (Primary Traits)

MORNING NEWS	EXPLANATION	EXPERT
Organized around *DETAILS*	**Organized around *LOGICAL STEPS***	**Organized around *CONCEPTUAL CATEGORIES***
• **Who/When/Where** - details regarding actors, time, and place • **Images** - detailed images regarding surroundings and events • **Temporality** - Events are ordered, linear in time. • **Newsworthiness** - text conveys importance or significance of event	• **Adequacy** - sufficiency of information regarding necessary steps to complete a task. • **Keywords** - signal a sequence (e.g., "next" "first" "second" "then")	• **Depth** - development within categories • **Breadth** - development across categories • **Keywords or Phrases** - signal a change of categories

An essential characteristic of the expert genre is a presence of cohesive conceptual categories or "chunks" of information, evaluated according to their depth and breadth. Another trait is textual signals, keywords or phrases, that signal that one is making a transition in categories (for example, the phrase "*Now*, I would like to talk about what crocodiles eat.").

The explanation genre, on the other hand, is characterized by an articulation of the logical or sequential steps through which an activity or task is accomplished. An explanation is evaluated primarily according to its thoroughness. Here, too, there are keywords or phrases, this time signaling one's place in a sequence or process (the words "next" or "then", for example). Finally, an essential characteristic of the narrative genre is the reporting of concrete and specific details as experienced through the eyes of a conscious actor. We evaluated this ability according to the writer's use of "who-what-when" information, visual imagery, use of time, as well as ability to communicate the personal significance of an event.

Table 2 shows the dimensions for Morning News. The other measures are similarly comprised. Also shown are scoring ranges for the different dimensions. In addition to the Primary Traits scale, each measure contained a Writing Conventions scale. This scale consisted of items intended to measure conventional writing skills such as spelling, paragraphing, capitalizing, and

so on. While the nature of Primary Traits varied from one measure to another, Writing

Convention items were identical for each of the three protocols. Common to all protocols was also a Reader Sensitivity Scale. This assessed the ability of the author to communicate his or her awareness of an audience, grab their attention, and reveal themselves as a person through the use of humor, dialogue, opinion, and so on. For each measure, the actual number of words produced, alone and then with help, were counted. Finally, the Expert and Explanation protocols included a

Table 2

Scoring Dimeñsions and Range For Morning News

Dimension	*Range*
• Holistic	0 - 3
• Primary Traits Total	0 - 18
• Rdr Sensitivity Total	0 - 9
• Writing Conventions Total	0 - 18

count of conceptual categories and logical steps, respectively.

In preparation for scoring, a team of graduate students was trained on the protocols until an adequate degree of inter-rater reliability was achieved (approximately 90-95%). All student texts were then scored separately by two different raters with any discrepancies resolved between them or, when necessary, by a third rater. The reliability and validity data that follow are based on a year-long pilot study of 137 students. We now move on to our first question: are LEAP's original measures both consistent and valid?

Consistency and Validity. Since these protocols are original, no evidence exists for their adequacy. If a measure is internally consistent, then its items will correlate with one another, suggesting that they are in fact measuring the same thing. Coefficient alpha, representing the average correlation between items, was computed based on a pilot study sample. These indicated good internal consistency for the primary traits scale, with alpha ranging from .79 to .89. Results for the Writing Convention scale, while somewhat lower, were still adequate with alpha for the three genre measures ranging from .64 to .74 (see appendix).

Evidence is also required for content validity. Content validity refers to whether a measure actually measures what it purports to measure. This is typically assessed by the tendency of a measure to correlate in expected ways with established measures of the same or similar

constructs. Validity for the convention scales for the three measures was assessed by correlating the total convention score for each measure with the "mechanics" score for "Shay's Story."

Shay's Story asks students to edit a short text containing various errors and omissions. The "mechanics" subscore indicates the number of technical problems, such as punctuation, capitalization, and spelling, that students are to identify and correct. Correlations between this subscale and the Writing Conventions scale of the original genre measures attained significance

Table 3: Correlations Between *Conventions (total)* and *SHAY (mechanical)* for LEAP's Three Original Genre Measures

All correlations $p < .001$

(Post)	SHAY (mechanical)	Expert Conventions	Explanation Conventions
SHAY (mechanical)	--		
Expert Conv	.52	--	
Explanation Conv	.59	.58	--
Morn. News Conv	.54	.63	.70

at the .01 level for each measure. This would argue that the convention scales are valid measures of students' awareness regarding the technical aspects of writing.

Establishing validity for the Primary Traits scales is a bit more tricky. Of course, face validity of the items themselves is important. Quantitatively, while our internal consistency estimates tell us that we are in fact measuring *something*, it would also be important to know that this "something" is, in fact, different for each of the three genres: expert, explanation, and narrative.

Table 4: Correlations Between *Writing Conventions* for LEAP's Three Original Genre Measures

All correlations $p < .001$

(Post)	Expert Total Conventions	Explanation Total Conventions	Morning News Total Conventions
Expert Total Conventions	--	.70	.67

Table 5: Correlations Between *Primary Traits* (Total) for 3 Original Genre Measures

All correlations $p < .001$

(Post)	Explanation Total Primary	Morning News Total Primary
Expert Total Primary	.49	.45

The number of students for each correlation ranges from 103 to 115

Table 4 shows the correlations between the Writing Convention scale of the three protocols. These are high.[2] On the other hand, Table 5 displays correlations between the Primary Traits scale of the three protocols. As you can see, the latter are comparatively lower. These suggest that while there is indeed a degree of commonality between the Primary Traits scales of each measure, neither are these scales identical.

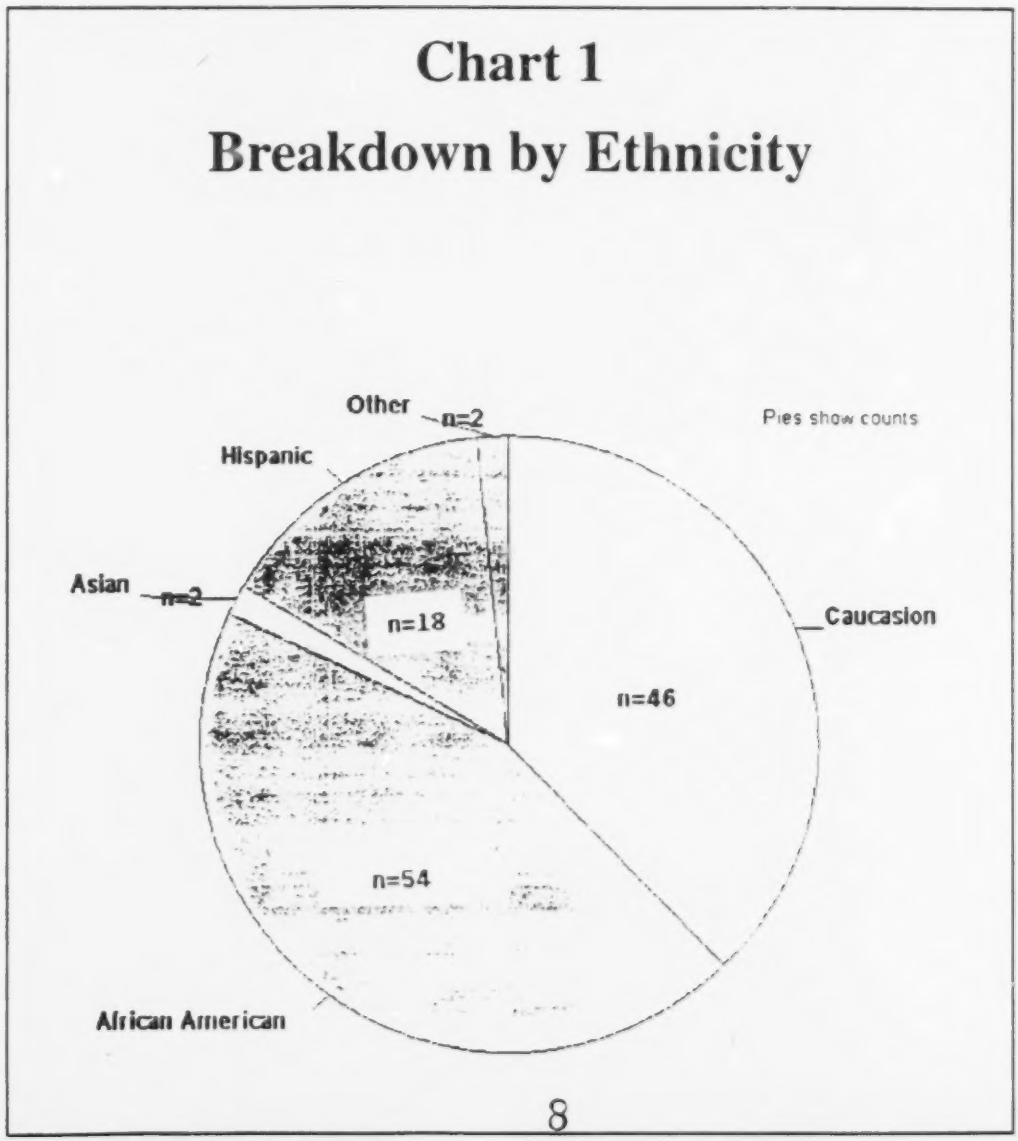
---

[2] Indeed, since the Writing Convention items are identical for each instrument, it is curious that these correlations are not higher.

We also anticipated a degree of correlation between dimensions *within* a measure. It is obvious that various aspects of the writing process are connected to one another. For example, one's ability to construct a text in *any* genre presumes a certain level of technical skill. It is that technical skill that allows one to produce a greater or lesser number of actual words, the words which one arranges into a well-structured text, perhaps a text that is particularly sensitive to its audience. For that reason, we expected a degree of inter-correlation between the different dimensions of our measures. Indeed, when computed separately for all three of our measures, the intercorrelations between their dimensions were very high, all significant, some as high as .94, none lower than .23, and the average somewhere around .50.

Demographics

Data from this point on are based on the year 1997-1998, subsequent to the pilot study. Our sample consisted of approximately 115 special education students; the number of students per



BEST COPY AVAILABLE

analysis varies. Although students ranged from kindergarten through sixth, the average was around third grade. The sample was predominately male, two to one (see appendix). Students were assessed during the fall and again the following spring.

The Ethnicity breakdown, primarily African-American and Caucasian, can be seen in Chart 1. There were three experimental groups. Students from four elementary LEAP classrooms, each consisting of both special education and general education students, comprised the LEAP Inclusion group. Students from six different elementary LEAP resource rooms comprised the LEAP Resource group while students from five non-LEAP resource rooms constituted the Comparison Group.

Does LEAP Have an Effect?

We now move onto our second general question: "Using the measures we have developed, does LEAP appear to have a statistical effect?" Because of the high correlation between the dimensions of the measures, we have chosen techniques designed to accommodate for the inter-relatedness of so many dependent variables while also controlling for Type I error. First, for each measure, we conducted a multivariate analysis of covariance, or MANCOVA. MANCOVA is simply an analysis of covariance in which several dependent variables are analyzed at the same time. Statistically significant analyses are followed up with a series of univariate analyses on each of the dependent variables. This was done in order to determine their level of contribution to the significant multivariate effect. From here, we turned to Discriminant Analyses (DA), a factoring technique that allows us to make even finer interpretations regarding the dimensions we are attempting to assess.[3]

By looking at the relative contributions of each of the dependent variables to a common factor, we can interpret this factor in practical terms. Finally, a univariate Analysis of Covariance by Group was performed using the students' loadings on the derived factor as the dependent variable. This final analysis allowed us to test for differences between the groups themselves on the earlier derived factor through more traditional post-hoc tests.

Owing to this session's time constraints, we will concern ourselves here with results from only one of the original measures: Morning News. In their Morning News assessment, students were directed to write about a personal experience or event much as they might *tell* about it during their class's Morning News activity. This activity is an approximately 15-20 minute interaction during which a student orally shares a personally significant event with the class. The class then

---

[3] Essentially, Discriminant Analysis (DA) is an extension of MANOVA that allows one to extract a common function, or what we will refer to here as composite or "factor," from a group of correlated variables and then determine a coefficient for each dependent variable that, similar to factor analysis, reflects its contribution to, or "loading" on, that factor (Nunnally, 1978).

Nunnally, J. C. (1978). *Psychometric theory*. (2nd ed.). New York: McGraw-Hill.

BEST COPY AVAILABLE

collaborates with the author and teacher to produce an edited text of the experience or event.

RESULTS

A single factor MANCOVA with five dependent variables and five covariates found a significant effect for Group among the six Morning News variables (Wilks' lamda = .002). There were no effects for gender, ethnicity, or type of disability.

[Table 6]

Because the numeric range of the different scales vary widely, in order to prevent confusion we present our results as *z*-scores. *Z*-scores are standardized scores with a mean of 0 and a standard deviation of 1. Thus, for example, a score of .50 would represent a score ½ a standard deviation above the mean. Also, where covariates were used, the means have been adjusted for the covariate.

Follow-up univariate analyses determined that Holistic, Primary Traits, and Total Words were the main dimensions primarily contributing to the multivariate effect. Contrasts revealed that, for the Holistic score, the adjusted mean for LEAP Inclusion exceeded that of LEAP Resource while LEAP Resource, on average, exceeded that of Comparisons. For Primary Traits *and* Student Words, while the adjusted means for neither LEAP Inclusion nor LEAP Resource significantly differed from one another, the adjusted means for both LEAP groups exceeded that of Comparisons. These results clearly suggest that in the area of narrative writing, word production, and the overall quality of narrative text, the means of those classrooms in which LEAP was implemented exceeded those of the comparison group.

While this was indeed encouraging, we were also troubled. Does this mean that LEAP had no impact in the area of technical skills, sensitivity to audience, and number of words students alone could produce? Perhaps this is where the attempt to break literacy down into discrete components falls short. It is intuitive that a certain synergism exists in any broadly comprehensive curriculum and, particularly under such conditions, that students' various types of thinking and acting are not independent of one another. As another type of follow-up, Discriminant Analysis (DA) was therefore used in order to create a composite variable, or "factor" that perhaps might capture this in a rough way. In essence, all six variables were "thrown into the hopper" and a single composite was created from the shared or "overlapping" variance existing between all six variables.

The result was that two composites or "factors" emerged.[4] The first of these accounted for approximately 70% of the shared variance while the second accounted for approximately 30%. By then looking at the Structure Matrix for these factors, we attempted to make a meaningful interpretation of what they might actually mean.

[Table 7]

---

[4] Two functions are the most that can be derived from a three-level variable, in this case Group.

The structure matrix contains what are often referred to as "loadings." These are simply the correlations between each dimension of the measure and the factor this dimension correlates with. In other words, they represent a particular dimension's level of contribution to Factor 1 or Factor 2. Pedhazur (1982) suggests that only loadings of .30 and above be included in one's interpretation.

With this in mind, we can see that relative to factor 2, factor 1 had higher loadings on Total Words, Total Student Words, Writing Conventions Total, and Reader Sensitivity. Total Primary Traits also loaded above .30 on this factor. One therefore might label Composite #1 *Basic Literacy Skill*. On the other hand, relative to Composite #1, Composite #2 revealed higher loadings on the Holistic and Primary Traits dimensions, with negligible contributions from the other 4 dimensions. Since Primary Traits are where the core aspects of a particular genre are assessed, and Holistic scores typically suffered if students' writing was *not* in the requested genre, composite #2 suggests an aspect of writing apart from basic technical skills, one reflecting qualities of the actual content or substance of the work as well as the writer's ability to create a well-structured, coherent text in a specific genre. Given the particular measure, Morning News, we might label this factor *Story Telling*. Now, our question was "How do the three groups compare on these two factors?"

To answer that question, we took each factor separately and computed an Analysis of Covariance on that factor with Group as the independent variable. Since pre-existing differences in students' literacy skills or narrative ability might obfuscate any genuine post-test differences, we used an appropriate covariate for each factor. For the Basic Literacy Skill factor, the covariate was students' pre-score for Total Writing Conventions. For the Genre Literacy factor, the covariate was students' pre-score on Total Primary Traits.

[Table 8]

The results of these analyses can be seen in Table 8. Group differences were significant for both factors. In post hoc tests, we found that for Factor 1, the mean for the Resource group was greater than that of both Inclusion and Comparison groups, which did not differ from one another. For Factor 2, the mean for the Inclusion group was greater than the means of both Resource and Inclusion groups, which did not differ from one another. Our interpretation of these factors was helped by the creation of a scatterplot showing the factor means for each group plotted simultaneously on each of the factors.

[Chart 2]

Factor 1 - Basic Literacy Skill. Looking only at the X-axis, that is, Factor 1 or the *Basic Literacy Skill* factor, an ellipsis surrounds the two groups that failed to differ from each other on the post hoc tests. Where our prior MANCOVA showed no statistically significant differences between the three groups on the Writing Conventions scale, it is interesting that here the mean for the LEAP Resource group was found to exceed the means of the other two.

Why the discrepancy? Perhaps it can be attributed to qualitative differences between the Writing Conventions scale and the derived Factor. Recall that, in addition to the Writing Conventions scale, the dimensions loading most heavily on Factor 1 were the number of actual written words as well as Reader Sensitivity. The important element in Reader Sensitivity, as we've defined it, is

the author's ability to "speak" or "interact" through his or her writing, to reveal aspects of his or her personality through humor or the sharing of opinions. Further refining our interpretation, perhaps what we are seeing is how self, basic literacy skills, and the ability to produce a written story ultimately coalesce. Intuitively, we all know they are not independent of one another, except in the abstract.

Looking at the group means themselves for the General Literacy Skill factor, it may be that LEAP Resource classrooms, relative to Comparison resource rooms, may somehow provide special education students with greater opportunities for this integration of self, skill, and story. The finding that Comparison and LEAP Inclusion groups did not differ on this factor is troublesome. It may point to important differences in the way LEAP was implemented within inclusion settings versus the resource rooms. Alternatively, it may indicate differences between these two contexts in how special education students were able to assimilate the curriculum.

[Chart 3]

We look now at the Y-axis, or the *Story Telling* factor (only the ellipses change between charts 2 and 3). Here, as we said, the mean for the LEAP Inclusion group exceeded the means for the other two groups. Recall that on this factor, both Holistic and Primary Traits (those dimensions of the protocol assessing narrative quality) loaded quite heavily with almost no contribution from the other dimensions. Unfortunately, since our design did not include a non-LEAP Inclusion group, we are prevented from interpreting Factor 2 at the level of LEAP. However, we can look at the results in terms of Resource versus Inclusion settings. At this level, LEAP Resource and Comparisons, both of them resource room conditions, did not differ statistically.

This raises compelling questions. For example, what might be unique about Inclusion classrooms that would tend to facilitate special education students' story telling under those conditions? Inclusions rooms were generally larger and, unlike Resource and Comparison groups, were not pull-out programs. Has more sustained proximity to general education students positively impacted the development of special education students' ability to write a well-structured narrative text? Again, the absence of a non-LEAP inclusion group prevents us from exploring this hypothesis. However, it is possible that our ethnographic data will be able to eventually shine some light on this question.

Summary

Looking to the future, our next phase of quantitative work will involve attempts to discover more specifically the exact areas in which students appear to respond, or not respond, to the LEAP approach. For example, what are the *individual* primary traits or technical skills that appear to be affected? Since this is a multi-year (although not longitudinal) study, with some teachers coming on board at different times, and some involved with various forms of this curriculum for almost a decade, what is the effect on students of teachers' tenure with the project? Also, since we did involve inclusion classrooms, general education students were in fact a part of this project, although their data was not presented here. Future analyses will therefore begin to explore the responses of general education students to the LEAP curriculum.

In summary, we began by asking two general questions. First, do LEAP's original genre measures appear to be both consistent and valid? And second, among our three experimental

groups, does the LEAP curriculum appear to have a statistical effect? Based on the evidence presented here, we believe we can answer both of these questions in the affirmative.

Appendix

LEAP 1997-1998

Internal Consistency Estimates for Original Measures

	Explanation		Morning News		Expert	
	Traits	*Conv*	*Traits*	*Conven*	*Traits*	*Conven*
PRE	.79	.74	.88	.69	.89	.74
POST	.83	.70	.86	.64	.87	.70

GRADE and GENDER

		#
GRADE	Kindergarten	3
	First	14
	Second	27
	Third	46
	Fourth	20
	Fifth	23
	Sixth	4
GENDER	Male	97
	Female	46

Table 6

Morning News

Univariate Analyses of Dimensions by Group (Inclusion, Resource, Comparison)

Standardized *z*-scores (Mean = 0, SD = 1)

df = 2/106

			Adjusted Means (*z*-scores)			
Variable	***F*-ratio**	***p*-value**	**Inclus**	**Res**	**Comp**	**Significant Contrasts [1]**
Holistic	**7.52**	**.001**	**.49**	**.02**	**-.43**	**Inclusion > Resource > Comparison ($p < .04$)**
Total Primary Traits	**6.84**	**.002**	**.43**	**.06**	**-.43**	**Inclusion Resource > Comparison ($p < .02$)**
Reader Sensitivity	**1.07**	**.347**	**--**	**--**	**--**	**--**
Writing Conventions	**1.28**	**.284**	**--**	**--**	**--**	**--**
Total Words	**4.18**	**.018**	**.00**	**.13**	**-.24**	**Inclusion Resource > Comparison ($p < .00$)**
Student Words	**2.14**	**.122**	**--**	**--**	**--**	**--**

[1] A line under two groups for a given variable indicates that the groups do not differ at the .05 level with respect to that variable.

BEST COPY AVAILABLE

Table 7

Morning Message

Factors: Percent of Common Variance Accounted For

	% of variance
Factor 1	**69.7**
Factor 2	**30.3**

Table 8

Morning News

Univariate Analyses of Factors by Group

Standardized *z*-scores (Mean = 0, SD = 1)

df = 2/114

			Adjusted Means (*z*-scores)			
Variable	***F*-ratio**	***p*-value**	**Inclus**	**Res**	**Comp**	**Significant Contrasts [1]**
Factor 1 * ***(Skills)***	6.78	.002	-.30	.33	-.34	Resource > Inclusion Comparison ($p < .01$)
Factor 2 ** ***(Story Telling)***	5.36	.006	.46	.00	-.36	Inclusion > Resource Comparison ($p < .05$)

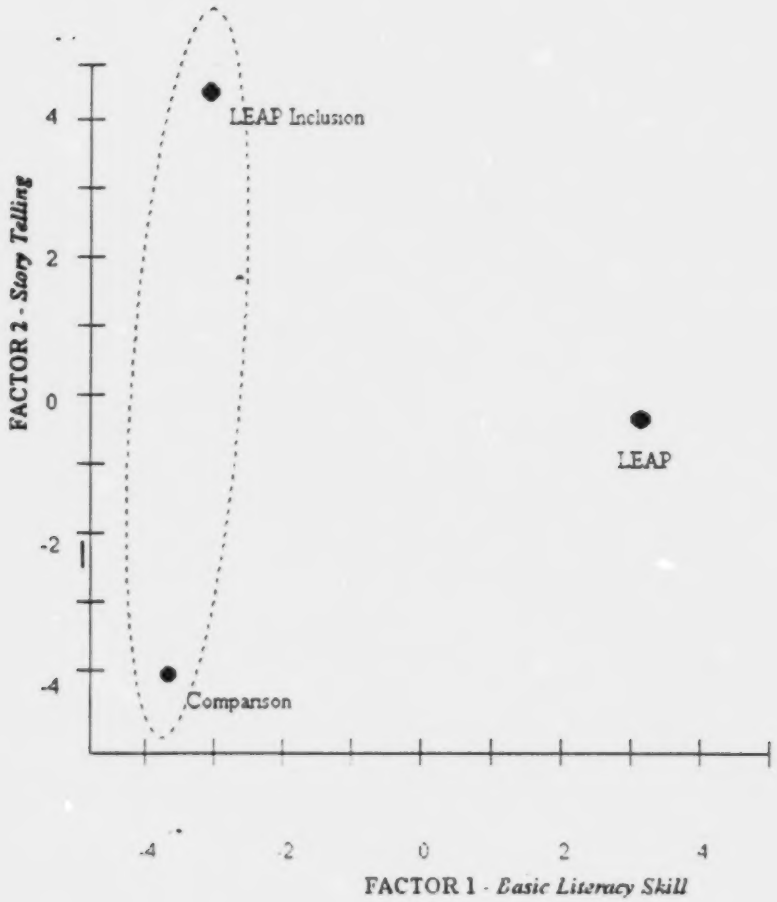
[1] A line under two groups for a given variable indicates that the groups do not differ at the .05 level with respect to that variable.

* The covariate for Factor 1 was the pre-test score for Writing Conventions (Total). In the analysis above, the effect of this covariate was significant ($p = .006$).

** The covariate for Factor 2 was the pre-test score for Primary Traits (Total). In the analysis above, the effect of this covariate was significant ($p = .008$).

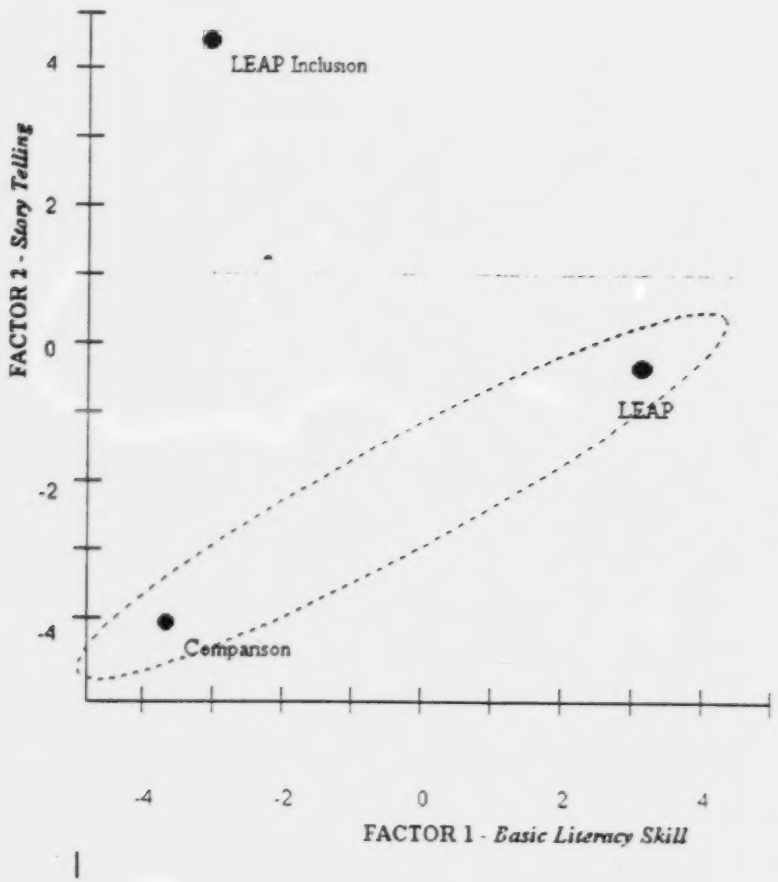
BEST COPY AVAILABLE

Chart 2



BEST COPY AVAILABLE

Chart 3


LEAP Inclusion
LEAP
Comparison
4
2
0
-2
-4
-4
-2
0
2
4
FACTOR 1 - Story Telling
FACTOR 1 - Basic Literacy Skill